JOURNAL OF COMPLEXITY 7, 35–57 (1991)

Noisy Information for Linear Problems in the Asymptotic Setting*

B. Z. KACEWICZ AND L. PLASKOTA

Institute of Informatics, University of Warsaw, 00-901 Warsaw, Poland

Received October 26, 1989

The asymptotic behavior of algorithms for solving linear problems is studied. Available information about the problem is assumed to be corrupted by bounded noise. We show how the asymptotic behavior is related to the worst case performance of algorithms. Algorithms and information that enjoy optimal convergence properties are obtained. Under some assumptions, optimal information turns out to be nonadaptive. © 1991 Academic Press, Inc.

1. INTRODUCTION

In this paper, we deal with the asymptotic behavior of algorithms for solving numerical problems. Such algorithms are based on some *information* about the problem, which is usually collected by computing (or measuring) the values of certain functionals. A typical example of information for problems such as function approximation or integration is given by function values at some points. In most cases information is corrupted by *noise* which may be caused by computation, measuring, or round-off errors. Even if exact information is available, it is sometimes impossible to store it because of memory restrictions. The exact computation of information functionals may also be far more expensive than their approximate evaluation.

For these reasons, we often use noisy information, which has been studied from different points of view by numerous authors, including numerical analysts, statisticians, and engineers. In many papers, noisy information is considered in the *worst* or *average case* settings. Results in

* Invited Paper.

0885-064X/91 $3.00

the first setting, in which one is interested in the worst performance of an algorithm and the information noise is assumed to be bounded, are briefly reported in Section 3. For details the reader is referred to the papers of Micchelli and Rivlin (1977), Melkman and Micchelli (1979), Lee *et al.* (1987), Marchuk and Osipenko (1975), and Kacewicz *et al.* (1986).

In the average case setting, in which we look at the average performance of an algorithm, information corrupted by stochastic noise was analyzed by Kadane *et al.* (1988) and Plaskota (1990). Using a statistical point of view, stochastic noise was studied in the papers of Kimeldorf and Wahba (1970), Wahba (1975, 1984), and Chaloner (1984).

In the present paper we consider noisy information in the *asymptotic* setting. We aim to find an algorithm that converges to the solution as fast as possible, when the number of information functionals goes to infinity (see Section 2). This setting has been previously studied for exact information. The first paper on this subject by is Trojan (1983), who showed a connection between the asymptotic and worst case settings for linear problems. Kacewicz (1987a) showed a similar connection for a class of nonlinear problems. A relation between the asymptotic and average case settings for linear problems was established by Wasilkowski and Woźniakowski (1987). Asymptotically optimal algorithms for evaluating the global maximum of a function were studied by Plaskota (1989) and for zero finding by Sikorski and Trojan (1987).

The asymptotic behavior of noisy information was considered in the unpublished paper of Trojan (1985). Roughly speaking, his goal was to determine the maximal accuracy of an approximation given by an (usually nonconvergent) algorithm. Such an approach is different from ours. In particular, the results of Trojan (1985) cannot be reduced to those of Trojan (1983) if information is exact.

We now describe the framework of the present paper. The problem being solved is linear. Information is corrupted by bounded noise which may tend to zero, as the number of information functionals goes to infinity. An example of such a situation is given in Example 2.1. Even though information is not exact, there may exist a convergent algorithm. If the noise is sufficiently small, the rate of convergence even may be the same as that for exact information; see the discussion in Section 5.

In order to study the influence of a noise on the rate of convergence, we show a close link between the asymptotic and worst case settings. Our approach is a generalization of that of Trojan (1983) to the case of noisy information. In Section 4 we show that, for given information, the optimal speed of convergence is determined by the diameter of information, which characterizes information in the worst case setting. The optimal speed of convergence is achieved by a spline algorithm. Furthermore, in Section 5 we show how to select optimal information functionals. This problem can be reduced essentially to finding such functionals in the worst case setting.

Another question considered in this paper is whether adaptive information is more powerful than nonadaptive information. It is known that adaption does not help when information is exact in the worst case, average case, and asymptotic settings; see, e.g., Traub *et al.* (1988, pp. 57–65, 236–247, 391–392) and Trojan (1983). The same holds for noisy information in the worst and average case settings; see Traub *et al.* (1988, p. 437), Kadane *et al.* (1988), and Plaskota (1990). We prove in Section 6 that adaption does not help (under some assumptions) in the case of noisy information in the asymptotic setting.

In Sections 2–6 we assume that the information values available during the computation process could change from step to step. In Section 7 we study a different (perhaps more natural) model, requiring perturbed information to be given as an *infinite* sequence. The main difficulty in such a model is establishing a sharp lower bound on the error. Such a bound, corresponding to that given in Section 6, is presented in Theorem 7.1.

In this paper, we have shown how to analyze the error of algorithms in the asymptotic setting when information is given with bounded noise. Using these results, one can study the (asymptotic) *ε-complexity* of linear problems, i.e., the minimal cost of computing an ε-approximation to the solution. To reduce the length of this paper, we do not develop this topic now. Complexity results will be presented in the future.

2. Formulation of the Problem

Let S: $F \to G$ be a linear continuous operator, where F is a Banach space and G is a normed space. We call S the *solution operator*. Our aim is to approximate an element Sf, called the *solution*, for $f \in F$. We assume that the element f is not known; we can only find some information about f. By (exact) information about f we mean a sequence

$$Nf = [L_1(f), L_2(f), \ldots, L_n(f), \ldots], \tag{2.1}$$

where L_i: $F \to \mathbb{R}$ are linear continuous functionals, $i = 1, 2, \ldots$. That is, we gain information about f by collecting the values of linear functionals at f. The mapping N: $F \to \mathbb{R}^\infty$ defined by (2.1) is called an *information operator* (or simply information). We denote the n first evaluations in (2.1) by $N^n(f)$; i.e.,

$$N^n(f) = [L_1(f), \ldots, L_n(f)]. \tag{2.1a}$$

In what follows, for a sequence $a = [a_1, a_2, \ldots]$, we let $a^n = [a_1, \ldots, a_n]$ denote the vector defined by the first n components of a.

The crucial assumption of this paper is that the evaluations of $L_i(f)$ are corrupted by noise. That is, instead of knowing the exact numbers $L_i(f)$ we know only their perturbed values z_i. For instance, we may know numbers z_i such that

$$|z_i - L_i(f)| \leq \Delta_i, \qquad i = 1, 2, \ldots, \tag{a}$$

for some $\Delta_i \geq 0$. Or, for each n we may know a vector $z^n \in \mathbb{R}^n$ such that

$$\|z^n - N^n f\| \leq \Delta_n, \tag{b}$$

where $\|\cdot\|$ is a norm on $\mathbb{R}^n$, and $\Delta_n \geq 0$.

To deal with the general case of perturbed information we assume that $\mathbb{R}^n$ is equipped with a functional $\|\cdot\|_n$: $\mathbb{R}^n \to [0, +\infty]$, which enjoys the properties of a norm (with the convention that $0 \cdot \infty = \infty \cdot 0 = 0$). That is, $\|\cdot\|_n$ is a norm on the subspace $\{x \in \mathbb{R}^n : \|x\|_n < +\infty\}$. We call $\|\cdot\|_n$ an *extended norm*. We also assume that the extended norm is independent of N and f.

By *nth noisy (perturbed) information* about f corresponding to the information operator N we mean a vector z^n such that

$$\|z^n - N^n f\|_n \leq 1. \tag{2.2}$$

Note that z^n is equal to exact information $N^n f$ if $\|x\|_n = +\infty$ for $x \neq 0$ and $\|0\|_n = 0$. Exact information is thus a special case of our considerations.

In example (a) we have

$$\|x\|_n = \max_{1 \leq i \leq n} |x_i|/\Delta_i, \tag{a1}$$

while in (b)

$$\|x\|_n = \|x\|/\Delta_n, \tag{b1}$$

where $x = [x_1, \ldots, x_n]$, with the convection that $a/0 = +\infty$ for $a > 0$ and $0/0 = 0$.

The solution Sf for $f \in F$ is approximated by an algorithm based on noisy information z^n. More specifically, by an *algorithm* we mean a sequence of mappings $\Phi = \{\Phi^n\}$, where Φ^n: $\mathbb{R}^n \to G$. The *nth error* of an algorithm Φ (with information N) at f is defined as

$$e_n(\Phi, N, f) = \sup\{\|Sf - \Phi^n(z^n)\| : \|z^n - N^n f\|_n \leq 1\}. \tag{2.3}$$

Note that, since we do not know what noisy information z^n we have to deal with, we define the error for the worst z^n consistent with exact information $N^n f$.

Note that the error is defined for each element f individually. We are interested in the asymptotic behavior of $e_n(\Phi, N, f)$ as $n \to +\infty$. We wish to define Φ in such a way that the convergence is as fast as possible. Such a setting is called *asymptotic*.

We now give examples of possible sources of an information noise. For instance, the numbers z_i may be results of *measurements* of the exact values $L_i(f)$. We deal then with the case (a), where Δ_i are the measurement errors.

The information noise may also be a result of previous stages of computation. Consider the following example.

EXAMPLE 2.1. Let $F = G = C([0, 1])$ with the maximum norm. Let $Sf = f$ and

$$Nf = [f(x_1), f(x_2), f(x_3), \ldots],$$

where $x_i \in [0, 1]$. That is, we wish to uniformly approximate a function f knowing its approximate values at certain points. Suppose that f is the solution of an initial value problem

$$\begin{aligned} f'(x) &= g(x, f(x)), \qquad x \in [0, 1], \\ f(0) &= 1, \end{aligned} \tag{2.4}$$

where g: $[0, 1] \times \mathbb{R} \to \mathbb{R}$ belongs to a class of sufficiently regular functions with uniformly bounded derivatives.

We assume that the values of f can be computed only via solving (2.4) (so that they are not exact), and that we are not involved in their computation. For instance, we do not know what information about the function g or what method was used to solve (2.4). Therefore, the problem cannot be reduced to that of approximating f on the basis of certain exact information about g; for such an analysis see Kacewicz (1987b). We only know information consisting of nonexact values of f and an upper bound on the noise.

What upper bounds can we expect to have? Suppose that we compute f at the points x_i chosen such that

$$x_1 = 0, \qquad x_2 = 1, \qquad x_{2^{i-1}+2+j} = 2^{-1} + j \cdot 2^{-(i-1)},$$

where $j = 0, 1, \ldots, 2^{i-1} - 1$, $i = 1, 2, \ldots$. Information $N^{2^i+1}f$, $i = 0, 1, \ldots$, is thus given by the values of f at $2^i + 1$ equidistant points from

$[0, 1]$. Suppose that a pth order method ($p \geq 1$) for solving (2.4) is used to approximate $f(x_i)$, $i = 1, 2, \ldots, n$. We then get approximations z_i such that

$$\max_{1 \leq i \leq n} |z_i - f(x_i)| = O(n^{-p}), \qquad n \to +\infty.$$

For many methods the constant in the O-notation is independent of f, if functions g have uniformly bounded derivatives. We then have the situation described in (b), with $\Delta_n = O(n^{-p})$. As we see, in this case $\Delta_n \to 0$ as $n \to +\infty$.

3. Noisy Information in the Worst Case Setting

We briefly recall some known result concerning noisy information in the *worst case setting*. In this setting we want the error to be small for the worst element f from some class $F_0 \subset F$. Let N be an information operator (2.1). The *nth worst case error* of an algorithm Φ using perturbed information corresponding to N is defined by

$$e_n^{\text{wor}}(\Phi, N) = \sup_{f \in F_0} \sup_{\|z^n - N^n f\|_n \leq 1} \|Sf - \Phi^n(z^n)\|;$$

see (2.3) for a comparison.

This setting was considered in a number of papers. For instance, Micchelli and Rivlin (1977) obtained a result, which in our notation reads as follows:

THEOREM 3.1. *Let F_0 be convex and balanced. Then*

$$\inf_{\Phi} e_n^{\text{wor}}(\Phi, N) = \alpha \cdot d_n(N),$$

where $d_n(N) = 2 \cdot \sup\{\|Sh\| : h \in F_0, \|N^n h\|_n \leq 1\}$ *and* $\frac{1}{2} \leq \alpha \leq 1$.

This theorem allows us to express the minimal worst case error independently of a particular algorithm by the quantity $d_n(N)$, which only depends on the information N. We call $d_n(N)$ the *nth diameter of information N*. As we see in the next sections, the diameter of information also plays an important role in the asymptotic setting. To give examples of the worst case approach we recall the following two results.

Let F be the space of functions f: $[0, 1] \to \mathbb{R}$ which have absolutely continuous $(r - 1)$st derivatives, $r \geq 1$, and let

$$F_0 = \{f \in F : \|f^{(r)}\|_\infty \leq 1\}.$$

Let $N^n f$ consist of function evaluations at n points; i.e.,

$$N^n f = [f(x_1), \ldots, f(x_n)].$$

We assume that the computed vector $z^n = [z_1, \ldots, z_n]$ satisfies

$$\|z^n - N^n f\| \leq \Delta_n,$$

where $\|x\| = (\sum_{i=1}^n |x_i|^p)^{1/p}$, $1 \leq p \leq +\infty$, $x = [x_1, \ldots, x_n]$. The extended norm in $\mathbb{R}^n$ is thus given by $\|x\|_n = \|x\|/\Delta_n$.

Now let $G = \mathbb{R}$ and $Sf = f(x)$ for some $x \in [0, 1]$; i.e., we want to recover the value of the function f at x. Marchuk and Osipenko (1975) showed that for $n = r$ the Lagrange interpolation gives the minimal error equal to $\frac{1}{2}d_n(N)$, and

$$d_n(N) = 2\,\frac{|w_n(x)|}{n!} + 2\Delta n\|\overline{P}(x)\|_q,$$

where $1/p + 1/q = 1$, $w_n(x) = \prod_{i=1}^n (x - x_i)$, $\overline{P}(x) = [P_1(x), \ldots, P_n(x)]$, and $P_j = \prod_{i\neq j} (x - x_j)/(x_i - x_j)$.

As the second example consider the function approximation problem, $Sf = f$, with $G = C([0, 1])$. This problem was recently studied by Lee *et al.* (1987). They showed that for $p = +\infty$ we have

$$\inf_N d_n(N) = C_r n^{-r} + D_r\Delta_n,$$

where

$$(K_r/\pi^r)(1 + o(1)) \leq C_r \leq (2/(\pi r))^{1/2}2^{-(r-1)}(1 + O(1/r)),$$
$$1 \leq D_r \leq 2(r/(2\pi))^{1/2}(1 + O(1/r)),$$

and K_r is the rth Favard constant. The best approximation is given by the piecewise Lagrange polynomial based on n equispaced points.

As we see, in both examples the diameter of information is influenced by the additional additive term which is proportional to the noise.

4. Asymptotic Setting

In this section we study the asymptotic behavior of algorithms using noisy information. Recall that the error of an algorithm Φ is defined by (2.3). We provide upper and lower bounds on the speed of convergence of the sequence of errors $e_n(\Phi, N, f)$, for $f \in F$. These bounds are given in terms of the nth diameter of information $d_n(N)$ defined in Theorem 3.1, with F_0 being the unit ball in F. That is, we now have

$$d_n(N) = 2 \cdot \sup\{\|Sh\| : h \in F, \|h\| \leq 1, \|N^n h\|_n \leq 1\}.$$

We first define the ρ-spline algorithm and next show that it enjoys the best convergence properties. Optimality properties of spline algorithms are known for many problems and settings; see, e.g., Micchelli and Rivlin (1977), Melkman and Micchelli (1979), Traub *et al.* (1988, pp. 95–101, 226–233), Trojan (1983), and Kacewicz (1987a). The following definition and theorem are generalizations of those for exact information (Trojan, 1983). Let N be an information operator. For $\rho > 1$ and $z \in \mathbb{R}^n$ being perturbed information for some f, let an element $\sigma^n z \in F$ be such that

(i) $\|N^n \sigma^n z - z\|_n \le 1$,
(ii) $\|\sigma^n z\| \le \rho \cdot \inf\{\|f\| : \|N^n f - z\|_n \le 1\}$.

Note that such an element exists. Due to (i) and (ii), it interpolates the data and satisfies almost minimal norm properties. Define the sequence of approximations to Sf by

$$\Phi_\rho^n(z) = S\sigma^n z, \qquad n = 1, 2, \ldots .$$

The algorithm $\Phi_\rho = \{\Phi_\rho^n\}$ is called the *p-spline algorithm*. We have

THEOREM 4.1. *For any $f \in F$ and $n = 1, 2, \ldots$,*

$$e_n(\Phi_\rho, N, f) \le K(f) \cdot d_n(N),$$

where $K(f) = \max\{1, ((1 + \rho)/2)\|f\|\}$.

Proof. Let $f \in F$, and let $z \in \mathbb{R}^n$ be such that $\|z - N^n f\|_n \le 1$. Then

$$\|N^n(f - \sigma^n z)\|_n \le \|N^n f - z\|_n + \|N^n \sigma^n z - z\|_n \le 2.$$

Hence, if $\|f - \sigma^n z\| \le 2$ then

$$\|Sf - \Phi_\rho^n(z)\| = \|S(f - \sigma^n z)\| \le d_n(N).$$

On the other hand, if $\|f - \sigma^n z\| > 2$ then we have

$$\begin{aligned} \|Sf - \Phi_\rho^n(z)\| &= \|S(f - \sigma^n z)\| \\ &= \|f - \sigma^n z\| \cdot \left\|S\left(\frac{f - \sigma^n z}{\|f - \sigma^n z\|}\right)\right\| \\ &\le \frac{1}{2}(\|f\| + \|\sigma^n z\|) \cdot d_n(N) \le \frac{1 + \rho}{2}\|f\| \cdot d_n(N). \end{aligned}$$

Hence, in both cases

$$\|Sf - \Phi_\rho^n(z)\| \le \max\left\{1, \frac{1+\rho}{2}\|f\|\right\} \cdot d_n(N).$$

Since z is arbitrary perturbed information for f, we get the desired bound. ■

The ρ-spline algorithm converges to the solution at least as fast as the sequence of the nth diameters. We now show that there is no algorithm with a speed of convergence essentially better than that of $\{d_n(N)\}$.

Let us assume that the extended norms satisfy the following condition:

(AS.1). If $\|[x_1, \ldots, x_n, x_{n+1}]\|_{n+1} < +\infty$ then $\|[x_1, \ldots, x_n]\|_n < +\infty$,

$\forall n, \forall x_i \in \mathbb{R}, i = 1, 2, \ldots, n+1$.

The condition (AS.1) holds for all the extended norms of interest. For instance, it is satisfied if $\|\cdot\|_n$ is a norm for all n. It also holds for the extended norm which defines exact information and for those defined by (a1) as well as by (b1) with the condition that if $\Delta_k = 0$ then $\Delta_m = 0$, $\forall m > k$.

The following theorem provides a lower bound on the error $e_n(\Phi, N, f)$ of an arbitrary algorithm Φ.

THEOREM 4.2. *Let N be an information operator such that $d_n(N) > 0$, $n = 1, 2, \ldots$, and Φ an algorithm which uses perturbed information corresponding to N. Let $\{\delta_n\}$ be an arbitrary positive sequence converging to zero. If* (AS.1) *holds then the set*

$$A_1 = \left\{f \in F : \limsup_{n\to+\infty} \frac{e_n(\Phi, N, f)}{\delta_n \cdot d_n(N)} < +\infty\right\}$$

is boundary in F, i.e., it does not contain a ball.

Proof. For any positive sequence $\{\delta_n\}$, define the set $\tilde{A}_1 = \tilde{A}_1(\{\delta_n\})$ by

$$\tilde{A}_1 = \left\{f \in F : \lim_{n\to+\infty} \frac{e_n(\Phi, N, f)}{\delta_n \cdot d_n(N)} = 0\right\}.$$

Then the set $A_1 = A_1(\{\delta_n\})$ is contained in $\tilde{A}_1(\{\sqrt{\delta_n}\})$. Hence, to prove the theorem it suffices to show that for any positive sequence $\{\delta_n\}$ converging to zero the set $\tilde{A}_1$ is boundary.

Suppose on the contrary that the set $\tilde{A}_1$ contains a closed ball B with radius ε, where $0 < \varepsilon \le 1$. To show that this is not possible we construct by induction a sequence $\{f_k\}_{k=1}^{\infty}$ contained in B and a sequence of integers $n_0 < n_1 < n_2 < \cdots$.

Let f_1 be the center of B, and $n_0 = 0$. Suppose that for some $k \geq 1$ we have constructed $f_1, f_2, \ldots, f_k \in B$ and integers $n_0 < \cdots < n_{k-1}$ such that $\|f_{i+1} - f_i\| \leq (\varepsilon/2)^i$, $i = 1, 2, \ldots, k-1$. To define f_{k+1} and n_k we proceed as follows. Let $\varepsilon_k > 0$ be such that

$$\|Sf - Sf_k\| \leq \tfrac{1}{3} \cdot \|Sf_k - Sf_{k-1}\| \qquad \text{for} \qquad \|f - f_k\| \leq \varepsilon_k$$

(in the case $k = 1$ we set $\varepsilon_1 > 0$ to be such that $\|Sf - Sf_1\| \leq \frac{1}{3}$ for $\|f - f_1\| \leq \varepsilon_1$).

The functional $\|\cdot\|_n$ is continuous on the subspace $\{x \in \mathbb{R}^n : \|x\|_n < +\infty\}$. From this and the condition (AS.1) it follows that we can choose (for $k \geq 2$) a positive number r_k for which it holds that

$$\text{if } \|N^{n_{k-1}}h\|_{n_{k-1}} < +\infty \text{ and } \|h\| \leq r_k$$
$$\text{then } \|N^{n_i}h\|_{n_i} \leq 2^{-k}, \qquad i = 1, 2, \ldots, k-1$$

(for $k = 1$ we set $r_1 = +\infty$).

Since $f_k \in \tilde{A}_1$, there exists an integer $n_k > n_{k-1}$ such that

$$\delta_{n_k} \leq \min\{\varepsilon_k, r_k, (\varepsilon/2)^k\},$$

and

$$\|Sf_k - \Phi^{n_k}(N^{n_k}f_k)\| \leq \tfrac{1}{10} \cdot \delta_{n_k} d_{n_k}(N).$$

From the definition of $d_{n_k}(N)$ there exists $h_k \in F$ such that

(i) $\|h_k\| \leq \delta_{n_k}$,
(ii) $\|N^{n_k}h_k\|_{n_k} \leq \delta_{n_k}$, and
(iii) $\|Sh_k\| \geq \frac{1}{4} \cdot \delta_{n_k} d_{n_k}(N)$.

We now set $f_{k+1} = f_k + h_k$. Since $\|f_{i+1} - f_i\| \leq (\varepsilon/2)^i$, $i = 1, 2, \ldots, k$, we have $\|f_{k+1} - f_1\| \leq \sum_{i=1}^{k} \|f_{i+1} - f_i\| \leq \varepsilon$, so that $f_{k+1} \in B$ and the construction of $\{f_k\}$ and $\{n_k\}$ is completed.

The sequence $\{f_k\}$ satisfies the Cauchy condition. Indeed, for any m, k, $m > k$, we have

$$\|f_m - f_k\| \leq \sum_{i=k}^{m-1} \|f_{i+1} - f_i\| \leq \sum_{i=k}^{m-1} (\varepsilon/2)^i \leq 2(\varepsilon/2)^k.$$

Thus, there exists $f^* \in B$ such that

$$f^* = \lim_{k \to +\infty} f_k.$$

We now show two properties of f^*. Observe first that since $\|f_{k+1} - f_k\| \leq \varepsilon_k$, we have $\|Sf_{k+1} - Sf_k\| \leq \frac{1}{3} \cdot \|Sf_k - Sf_{k-1}\|$, $k = 2, 3, \ldots$. This gives for $m > k$

$$\|Sf_m - Sf_k\| \geq \|Sf_{k+1} - Sf_k\| - \sum_{i=k+1}^{m-1} \|Sf_{i+1} - Sf_i\|$$

$$\geq \left(1 - \sum_{i=k+1}^{m-1} (\tfrac{1}{3})^{i-k}\right) \|Sf_{k+1} - Sf_k\| \geq \tfrac{1}{2} \cdot \|Sf_{k+1} - Sf_k\|.$$

By letting $m \to +\infty$ we get

$$\|Sf^* - Sf_k\| \geq \tfrac{1}{2} \cdot \|Sf_{k+1} - Sf_k\|, \qquad k = 1, 2, \ldots.$$

Second, since $\|N^{n_i} h_i\|_{n_i} < +\infty$, $\forall i$, due to (AS.1) we have that $\|N^{n_k} h_i\|_{n_k} < +\infty$, for $i \geq k$. Since also $\|h_i\| \leq r_i$, we have $\|N^{n_k} h_i\|_{n_k} \leq 2^{-i}$. Hence

$$\|N^{n_k} f_m - N^{n_k} f_k\|_{n_k} \leq \sum_{i=k}^{m-1} \|N^{n_k} h_i\|_{n_k}$$

$$\leq \sum_{i=k}^{m-1} 2^{-i} \leq 1, \qquad \forall m > k.$$

Since $\|\cdot\|_{n_k}$ is continuous on the closed subspace $\{x \in \mathbb{R}^{n_k} : \|x\|_{n_k} < +\infty\}$, we get

$$\|N^{n_k} f^* - N^{n_k} f_k\|_{n_k} \leq 1.$$

Thus, $N^{n_k} f_k$ may be treated as n_kth perturbed information for f^*.

Thus, we finally get

$$e_{n_k}(\Phi, N, f^*) \geq \|Sf^* - \Phi^{n_k}(N^{n_k} f_k)\| \geq \|Sf^* - Sf_k\|$$

$$- \|Sf_k - \Phi^{n_k}(N^{n_k} f_k)\|$$

$$\geq \tfrac{1}{2} \cdot \|Sf_{k+1} - Sf_k\| - \tfrac{1}{10} \cdot \delta_{n_k} d_{n_k}(N) \geq \tfrac{1}{40} \cdot \delta_{n_k} d_{n_k}(N),$$

for $k = 1, 2, \ldots$. This contradicts the fact that $f^* \in \tilde{A}_1$. The set $\tilde{A}_1$ is thus boundary and so is A_1, as claimed. ∎

Theorem 4.2 says that there is no algorithm converging to the solution as fast as the sequence $\{\delta_n \cdot d_n(N)\}$ (except for a boundary set of elements f). This together with Theorem 4.1 establishes almost optimal convergence properties of the ρ-spline algorithm Φ_ρ. Up to the sequence $\{\delta_n\}$, the optimal rate of convergence is given by the sequence $\{d_n(N)\}$ of the nth diameters of information N.

Although the sequence $\{\delta_n\}$ may converge to zero arbitrarily slowly, it cannot be neglected in the formulation of Theorem 4.2. Indeed, for the algorithm Φ_ρ we have $e_n(\Phi_\rho, N, f) = O(d_n(N))$, for all $f \in F$. Sometimes it is however possible to define an algorithm converging faster than $d_n(N)$. This means that not only A_1, but even the set $\tilde{A}_1$, which appears in the proof, may be not boundary if we neglect the sequence $\{\delta_n\}$.

EXAMPLE 4.1. Let $F = G$ be the Banach space of continuous functions f: $[0, 1] \to \mathbb{R}$, with the supremum norm. Let $Sf = f$ for $f \in F$. Consider the information operator

$$Nf = [f(x_1), f(x_2), \ldots, f(x_n), \ldots],$$

where the points x_i, $i = 1, 2, \ldots$, are given as in Example 2.1. Let noisy information z^n about f satisfy $\|z^n - N^n f\|_\infty \le \Delta_n$, $n = 1, 2, \ldots$, where $\{\Delta_n\}$ is a nonincreasing sequence converging to zero. It is not difficult to see in this case that $d_n(N) = 2$, for all n. On the other hand, consider the approximation of f by the piecewise linear function $h_n(f)$ which interpolates information z^n. Then, since f is continuous, $\{x_n\}$ is dense in $[0, 1]$, and $\Delta_n \to 0$, we have that $\|f - h_n(f)\|_\infty \to 0$ as $n \to +\infty$, for all f. However, convergence may be arbitrarily slow.

The next example shows that the assumption that F is a Banach space is important.

EXAMPLE 4.2. Let F be the space of functions f: $[0, 1] \to \mathbb{R}$ satisfying a Lipschitz condition, with the supremum norm. This is *not* a Banach space. Consider the space G, operators S, N, and noisy information as in Example 4.1. We have then again that $d_n(N) = 2$, $\forall n$. On the other hand, it is not difficult to check that $\|f - h_n(f)\|_\infty = O(\Delta_n + 1/n)$, $\forall f$, where the constant in the O-notation depends only on the Lipschitz constant for f. Hence, Theorem 4.2 does not hold. This shows that the assumption that F is a Banach space cannot be omitted in the formulation of Theorem 4.2.

5. OPTIMAL INFORMATION

We now show how the functionals L_i should be selected in order to maximize the speed of convergence.

Assume that the linear continuous information functionals can be chosen from some class Λ. Denote by $d(n, \Lambda)$ the nth minimal diameter with respect to Λ, i.e.,

$$d(n, \Lambda) = \inf d_n(N),$$

where the infimum is taken over all information operators (2.1) consisting of functionals from Λ. Assume also that $d(n, \Lambda) > 0$ for $n = 1, 2, \ldots$. We have

THEOREM 5.1. *Let N be an information operator (2.1) defined by functionals from Λ and Φ an algorithm which uses perturbed information corresponding to N. Let $\{\delta_n\}$ be an arbitrary positive sequence converging to zero. If* (AS.1) *holds then the set*

$$\left\{f \in F : \limsup_{n \to +\infty} \frac{e_n(\Phi, N, f)}{\delta_n \cdot d(n, \Lambda)} < +\infty\right\}$$

is boundary in F.

Proof. It follows immediately from Theorem 4.2 and the fact that

$$d(n, \Lambda) \leq d_n(N) \quad \blacksquare$$

The above theorem says that the sequence of errors $e_n(\Phi, N, f)$ cannot tend to zero essentially faster than the sequence $d(n, \Lambda)$ of the nth minimal diameters, except for a boundary set of elements f. An information operator N^* for which there exists an algorithm with the error proportional to $d(n, \Lambda)$ is called *optimal*.

How can we obtain optimal information? Assume that the extended norms satisfy the following assumption:

(AS.2). $\|[x_1, x_2, \ldots, x_n]\|_n \leq \min \{\|[x, x_1, \ldots, x_n]\|_{n+1}, \|[x_1, \ldots, x_n, x]\|_{n+1}\}$, for all n, x, and $x_1, \ldots, x_n$.

The condition (AS.2) holds for instance for the extended norms defined by (a1), as well as for those defined by (b1) with $\|\cdot\|$ being the L_p norm in $\mathbb{R}^n$, $1 \leq p \leq +\infty$, both with nonincreasing sequences $\{\Delta_i\}$. Note that (AS.2) implies (AS.1).

Let $\eta > 1$. For any positive integer n, let information N_n consist of functionals from Λ and be chosen such that

$$d_n(N_n) \leq \eta \cdot d(n, \Lambda).$$

Define

$$N^* = [N_1^1, N_2^2, N_4^4, \ldots, N_{2^k}^{2^k}, \ldots], \tag{5.1}$$

where N_n^n denotes the first n functionals of N_n. The following theorem yields that in many cases information N^* is optimal.

THEOREM 5.2. *Let Φ_ρ be the ρ-spline algorithm which uses perturbed information corresponding to N^*. If* (AS.2) *holds then for any $n \geq 1$ and $f \in F$ we have*

$$e_n(\Phi_\rho, N^*, f) \leq \eta \cdot K(f) \cdot d\left(\left\lceil \frac{n+1}{4} \right\rceil, \Lambda\right),$$

where $K(f)$ is given in Theorem 4.1.

Proof. For $n \geq 1$, let $k = k(n)$ be the greatest integer satisfying $n \geq \sum_{i=0}^{k} 2^i$. Then all the functionals from $N_{2^k}^{2^k}$ are contained in $(N^*)^n$. From the condition (AS.2) it follows that $\|(N^*)^n h\|_n \leq 1$ implies $\|N_{2^k}^{2^k} h\|_{2^k} \leq 1$. Consequently,

$$\begin{aligned} d_n(N^*) &= 2 \cdot \sup\{\|Sh\| : \|h\| \leq 1, \|(N^*)^n h\|_n \leq 1\} \\ &\leq 2 \cdot \sup\{\|Sh\| : \|h\| \leq 1, \|N_{2^k}^{2^k} h\|_{2^k} \leq 1\} = d_{2^k}(N_{2^k}). \end{aligned}$$

It is easy to see that the sequence $d(n, \Lambda)$ is nonincreasing. Since $2^k \geq \lceil (n+1)/4 \rceil$, using Theorem 4.1 we get

$$\begin{aligned} e_n(\Phi_\rho, N^*, f) &\leq K(f) \cdot d_n(N^*) \leq K(f) \cdot d_{2^k}(N_{2^k}) \\ &\leq K(f) \cdot \eta \cdot d(2^k, \Lambda) \leq K(f) \cdot \eta \cdot d\left(\left\lceil \frac{n+1}{4} \right\rceil, \Lambda\right), \end{aligned}$$

which completes the proof. ■

Hence, the error of the ρ-spline algorithm based on noisy information corresponding to N^* is proportional to $d(\lceil (n+1)/4 \rceil, \Lambda)$. Recall that Theorem 5.1 gives a general lower bound essentially equal to $d(n, \Lambda)$. Thus, if

$$d\left(\left\lceil \frac{n+1}{4} \right\rceil, \Lambda\right) = O(d(n, \Lambda)),$$

then Φ_ρ based on N^* has the best convergence properties among all algorithms that use perturbed information corresponding to an arbitrary information operator defined by functionals from the class Λ. Hence, in such a case information N^* given in (5.1) is optimal. This holds, for example, when $d(n, \Lambda)$ behaves as a polynomial in $1/n$.

We now apply these results to the approximation problem described in Example 2.1. We want to approximate the solution f of the initial value problem (2.4). We assume that $f \in F_0$, where

$$F_0 = \left\{ f \in C^{(r)}([0, 1]) : \sum_{i=0}^{r} \|f^{(i)}\|_\infty \leq 1 \right\},$$

and $r \geq 1$. Recall that the class Λ consists of functionals defined by the values of f at some points from $[0, 1]$. These values are (approximately) computed by a pth-order method for solving (2.4). The maximum norm error in computing n evaluations of f is then bounded by $\Delta_n = \Theta(n^{-p})$ (the constants in the Θ-notation are independent of f).

By applying the results of Lee *et al.* (1987) (see Section 3) we get that the minimal diameter of information for this problem is equal to

$$d(n, \Lambda) = \Theta(n^{-r} + n^{-p}), \qquad n \to +\infty.$$

In the asymptotic setting, Theorems 5.1 and 5.2 give (after minor modifications) that the optimal speed of convergence is also (essentially) of order $n^{-r} + n^{-p}$. Optimal information consists of evaluations of f at the points defined as in Example 2.1, and the optimal algorithm is given by piecewise Lagrange interpolation.

Note that for exact information (i.e., when $\Delta_n = 0$ for all n) the best rate of convergence is $\Theta(n^{-r})$. To preserve this speed of convergence in the presence of a noise, we have to compute optimal information with accuracy $\Delta_n = O(n^{-r})$. Hence, the problem (2.4) is to be solved using a method of order $p \geq r$.

We finally comment on methods for solving (2.4). It is known that for any method based on n values of the right-hand side function g (or its derivatives) the minimal achievable error Δ_n is of order $n^{-(r-1)}$ (see Kacewicz (1987b)). If such a method is used to compute information about f then the speed of convergence $O(n^{-r})$ cannot be achieved. The desired noise level $\Delta_n = \Theta(n^{-r})$ can be obtained by using some nonstandard (integral) information about g and nonstandard methods (see the same reference).

6. Adaptive Information

In the previous sections we have considered information defined by linear functionals which are given a priori, independently of f. Such information is called *nonadaptive*. In this section we study a more general class of information, allowing the functionals to be adaptively chosen for each f. More specifically, let $N: F \times \mathbb{R}^\infty \to \mathbb{R}^\infty$ be a mapping defined for $f \in F$ and $z \in \mathbb{R}^\infty$ by

$$N_z f = [L_1(f), L_2(f; z_1), \ldots, L_n(f; z_1, \ldots, z_{n-1}), \ldots], \quad (6.1)$$

where $z = [z_1, z_2, \ldots]$ and $L_i(\cdot; z_1, \ldots, z_{i-1})$, $i = 1, 2, \ldots$, are linear continuous functionals from the class Λ of permissible functionals. We call the mapping N an *adaptive information operator*. Observe that each information operator (2.1) is a special case of (6.1). For a fixed $z \in \mathbb{R}^\infty$, we denote by N_z the nonadaptive information operator, $N_z: F \to \mathbb{R}^\infty$, defined by (6.1).

A vector z^n is said to be *nth noisy information* about f corresponding to the adaptive information operator N if

$$\|z^n - N_z^n f\|_n \leq 1. \quad (6.2)$$

Observe that the functionals L_i in (6.1) are now chosen as functions of the perturbed values $z_1, \ldots, z_{i-1}$ of the previous functionals at f.

Let $\Phi = \{\Phi^n\}$ be an algorithm. For $f \in F$ the nth approximation to Sf is given by $\Phi^n(z^n)$, where $\|z^n - N_z^n f\|_n \leq 1$. Similarly to nonadaptive information, the *nth error* (at f) of an algorithm Φ using information operator N is defined by

$$e_n(\Phi, N, f) = \sup\{\|Sf - \Phi^n(z^n)\| : \|z^n - N_z^n f\|_n \leq 1\}.$$

We now give a lower bound on the error of an algorithm in cases where the information error is measured by a norm, or when information is exact. This bound is provided by the nth minimal diameter $d(n, \Lambda)$ of nonadaptive information. This leads to the conclusion that adaption is not more powerful than nonadaption.

Assume that $d(n, \Lambda) > 0$ for $n \geq 1$.

THEOREM 6.1. *Let N be an adaptive information operator and Φ be an algorithm that uses perturbed information corresponding to N. Let $\{\delta_n\}$ be an arbitrary positive sequence converging to zero. If the extended norms $\|\cdot\|_n$ are norms, or if they define exact information, then the set*

$$A_3 = \left\{f \in F : \limsup_{n \to +\infty} \frac{e_n(\Phi, N, f)}{\delta_n \cdot d(n, \Lambda)} < +\infty\right\}$$

is boundary in F.

Proof. We write N_{z^n} instead of N_z if only the first n components z^n of the sequence z are important and the remaining components do not enter into consideration. For $f \in F$ and $n \geq 1$, let

$$Z_n(f) = \{z^n \in \mathbb{R}^n : \|z^n - N_{z^n}^n f\|_n \leq 1\}.$$

It suffices to show that the set

$$\tilde{A}_3 = \left\{ f \in F : \forall \{z^n\}_{n=1}^{\infty} \text{ with } z^n \in Z_n(f) \text{ for } n \geq 1, \text{ it holds} \right.$$
$$\left. \lim_{n \to +\infty} \frac{\|Sf - \Phi^n(z^n)\|}{\delta_n \cdot d_n(N_{z^n})} = 0 \right\}$$

is boundary. We proceed by contradiction following the steps of the proof of Theorem 4.2. We construct sequences $\{f_k\}$, $\{n_k\}$ and an element $f^* \in \tilde{A}_3$ such that $z^{n_k} \in Z_{n_k}(f^*)$, where the components of z^{n_k} are defined by $z_i = L_i(f_k; z_1, \ldots, z_{i-1})$, $i = 1, 2, \ldots$. To go through the construction it is now not enough to assume only (AS.1), as we did in Theorem 4.2, since information is now adaptive. The only formal change with respect to the proof of Theorem 4.2 is that nonadaptive information N and N^{n_j} should be replaced by $N_{z^{n_j}}$ and $N^{n_j}_{z^{n_j}}$, respectively. In this way, we get that

$$\|Sf^* - \Phi^{n_k}(z^{n_k})\| \geq \tfrac{1}{40} \cdot \delta_{n_k} d_{n_k}(N_{z^{n_k}}),$$

for $k \geq 1$, which contradicts the fact that $f^* \in \tilde{A}_3$. ∎

It has been shown in Theorem 5.2 under the assumption (AS.2) that the ρ-spline algorithm Φ_ρ using perturbed information corresponding to the nonadaptive information operator N^* converges as fast as $d(\lceil n + 1)/4\rceil, \Lambda)$. In view of Theorem 6.1 we thus have

COROLLARY 6.1. *Let the extended norms $\|\cdot\|_n$ be as in Theorem 6.1, and let $d(n, \Lambda) > 0$ for $n \geq 1$. If (AS.2) holds and*

$$d\left(\left\lceil \frac{n+1}{4} \right\rceil, \Lambda\right) = O(d(n, \Lambda)),$$

then the ρ-spline algorithm Φ_ρ using perturbed information corresponding to the nonadaptive information operator N^* enjoys the best convergence properties among all algorithms based on perturbed information corresponding to an arbitrary (possibly adaptive) information operator (6.1).

That is, even though adaption is permitted (allowing us to adjust computations to a specific problem being solved), it does not improve speed of convergence.

7. A DIFFERENT MODEL OF NOISY INFORMATION

Up to now, noisy information z^n about an element f has been defined for each n separately. That is, the length and also all components of the vector z^n could change with n. In this section we consider another, perhaps more natural, approach. By *noisy information* about f corresponding to an adaptive information operator N we now mean an *infinite* sequence $z \in \mathbb{R}^\infty$ satisfying

$$\|z^n - N_z^n f\|_n \leq 1, \tag{7.1}$$

for all $n \geq 1$. Recall that z^n is a vector defined by the n first components of z.

Let $Z(f)$ be the set of $z \in \mathbb{R}^\infty$ for which (7.1) holds. Each element of $Z(f)$ is thus a sequence of perturbed evaluations of adaptive information functionals at f. It is now reasonable to study asymptotic behavior of the error

$$e_n(\Phi, N, f, z) = \|Sf - \Phi^n(z^n)\|, \tag{7.2}$$

for all $z \in Z(f), f \in F$.

Without loss of generality we assume that the extended norms $\|\cdot\|_n$ satisfy the following condition:

(AS.3). $\min\limits_{x \in \mathbb{R}} \|[x_1, \ldots, x_n, x]\|_{n+1} = \|[x_1, \ldots, x_n]\|_n$, for all $x_1, \ldots, x_n$ and $n \geq 1$.

This condition yields that if $z^{n+1} = [z_1, \ldots, z_n, z_{n+1}]$ is acceptable as a computed sequence for f (i.e., z^{n+1} belongs to the set $Z_{n+1}(f)$ defined in the proof of Theorem 6.1) then also $z^n = [z_1, \ldots, z_n]$ is acceptable for f. Furthermore, it yields that any vector $z^n \in Z_n(f)$ can be completed to an infinite acceptable sequence z, $z \in Z(f)$. The natural example of extended norms satisfying (AS.3) is provided by (a1).

We now explain why the assumption (AS.3) is not restrictive. We show that for any family of extended norms $\{\|\cdot\|_n\}$ there exists a family $\{\|\cdot\|_n'\}$ which satisfies (AS.3) and is equivalent (in some sense) to $\{\|\cdot\|_n\}$. Indeed, let $\|\cdot\|_*$ be an extended norm on $\mathbb{R}^\infty$ defined by

$$\|x\|_* = \sup_n \|x^n\|_n, \qquad \forall x \in \mathbb{R}^\infty,$$

and let

$$\|x\|_n' = \inf_{y \in \mathbb{R}^\infty} \|[x, y]\|_*, \qquad \forall x \in \mathbb{R}^n,$$

for $n \geq 1$. It is not difficult to see that $\|\cdot\|_n'$ are extended norms for which (AS.3) holds. Furthermore, it is possible to show that $\|\cdot\|_n$ satisfies (AS.3) for all n if and only if $\|\cdot\|_n = \|\cdot\|_n'$ for all n. Consequently, $\|\cdot\|_n' = (\|\cdot\|_n')'$.

Equivalence of the families $\{\|\cdot\|_n\}$ and $\{\|\cdot\|_n'\}$ is established in the following

LEMMA 7.1. *For $f \in F$ we have*

$$Z(f) = Z'(f),$$

where $Z'(f)$ is defined as $Z(f)$, with $\|\cdot\|_n$ replaced by $\|\cdot\|'_n$.

Proof. Let $z \in Z(f)$, and $x = z - N_z f$. Then $\|x^n\|_n \leq 1$, so that $\|x\|_* \leq 1$. Hence $\|z^n - N_z^n f\|'_n = \|x^n\|'_n \leq \|x\|'_* \leq 1$, for all $n \geq 1$, which gives $z \in Z'(f)$.

Now let $z \in Z'(f)$. Then for any $\varepsilon > 0$ and integer n there is $y \in \mathbb{R}^\infty$ such that

$$\|x^n\|_n \leq \|[x^n, y]\|_* \leq 1 + \varepsilon,$$

which yields that $z \in Z(f)$. ■

Hence, the extended norms $\|\cdot\|_n$ and $\|\cdot\|'_n$ lead to the same set of acceptable perturbed information. If (AS.3) does not hold for $\|\cdot\|_n$ then we replace it in the following considerations by $\|\cdot\|'_n$.

We now establish the asymptotic behavior of the error $e_n(\Phi, N, f, z)$, for $z \in Z(f)$. It is not difficult to see that for the ρ-spline algorithm Φ_ρ based on noisy information corresponding to an adaptive information operator N, we have

$$e_n(\Phi_\rho, N, f, z) \leq K(f) \cdot d_n(N_z), \qquad \forall z \in Z(f), f \in F$$

(compare with Theorem 4.1).

The fact that this bound cannot be improved, no matter what algorithm we consider, is less obvious. We show this in the following

THEOREM 7.1. *Let N be an adaptive information operator such that $d_n(N_z) > 0$ for all z and $n \geq 1$, and let Φ be an algorithm which uses perturbed information corresponding to N. Let $\{\delta_n\}$ be an arbitrary positive sequence converging to zero. Then the set*

$$A_4 = \left\{f \in F : \forall z \in Z(f) \limsup_{n \to +\infty} \frac{e_n(\Phi, N, f, z)}{\delta_n \cdot d_n(N_z)} < +\infty\right\}$$

is boundary in F.

Proof. As in the proof of Theorem 4.2, it suffices to show that the set

$$\tilde{A}_4 = \left\{f \in F : \forall z \in Z(f) \lim_{n \to +\infty} \frac{e_n(\Phi, N, f, z)}{\delta_n \cdot d_n(N_z)} = 0\right\}$$

is boundary.

Suppose on the contrary that $\tilde{A}_4$ contains a closed ball B with radius ε, where $0 < \varepsilon \leq 1$. We construct by induction a sequence $\{f_k\}_{k=1}^{\infty} \subset B$, a sequence of integers $n_0 < n_1 < n_2 < \cdots$, and $z^* = [z_1^*, z_2^*, \ldots] \in \mathbb{R}^{\infty}$. Let f_1 be the center of B and $n_0 = 0$. Suppose that for some $k \geq 1$ we have constructed elements $f_1, \ldots, f_k \in B$, integers $n_0 < n_1 < \cdots < n_{k-1}$, and (for $k \geq 2$) the first n_{k-1} elements $z_1^*, z_2^*, \ldots, z_{n_{k-1}}^*$ of z^*, such that

$$\|N_{z^*}^{n_i} f_{i+1} - (z^*)^{n_i}\|_{n_i} \leq \sum_{j=1}^{i} 2^{-j}$$

and

$$\|f_{i+1} - f_i\| \leq (\varepsilon/2)^i,$$

for $1 \leq i \leq k - 1$. From (AS.3) it follows that (for $k \geq 2$) there exist numbers x_i, for $i \geq n_{k-1} + 1$, such that the sequence $z = [z_1^*, z_2^*, \ldots, z_{n_{k-1}}^*, z_{n_{k-1}+1}, z_{n_{k-1}+2}, \ldots]$ with $z_i = L_i(f_k; z_1^*, \ldots, z_{n_{k-1}}^*, z_{n_{k-1}+1}, \ldots, z_{i-1}) + x_i$ satisfies

$$\|N_{z^*}^{n_{k-1}} f_k - (z^*)^{n_{k-1}}\|_{n_{k-1}} = \|N_z^i f_k - z^i\|_i, \qquad i \geq n_{k-1}$$

(for $k = 1$ we set $x_i = 0, i = 1, 2, \ldots$, so that $\|N_z^i f_1 - z^i\|_i = 0, \forall i$). Hence $z \in Z(f_k)$.

Choose $\varepsilon_k > 0$ such that $\|Sf - Sf_k\| \leq \frac{1}{3} \cdot \|Sf_k - Sf_{k-1}\|$ for $\|f - f_k\| \leq \varepsilon_k$ (for $k = 1$ we set ε_1 to be such that $\|Sf - Sf_1\| \leq \frac{1}{3}$ for $\|f - f_1\| \leq \varepsilon_1$).

Since $f_k \in B$, we can select an integer $n_k > n_{k-1}$ in such a way that

$$\delta_{n_k} \leq \min(\varepsilon_k, (\varepsilon/2)^k),$$

and

$$e_{n_k}(\Phi, N, f_k, z) \leq \tfrac{1}{10} \cdot \delta_{n_k} d_{n_k}(N_z).$$

From the definition of $d_{n_k}(N_z)$ there exists $h_k \in F$ such that

(i) $\|N_z^{n_k} h_k\|_{n_k} \leq \delta_{n_k}$,
(ii) $\|h_k\| \leq \delta_{n_k}$, and
(iii) $\|Sh_k\| \geq \frac{1}{4} \cdot \delta_{n_k} d_{n_k}(N_z)$.

We now define $f_{k+1} = f_k + h_k$ and $z_i^* = z_i$ for $n_{k-1} + 1 \leq i \leq n_k$. Note that for $k = 1$ we have

$$\|N_{z^*}^{n_1} f_2 - (z^*)^{n_1}\|_{n_1} = \|N_{z^*}^{n_1} h_1\|_{n_1} \leq \delta_{n_1} \leq \tfrac{1}{2},$$

which motivates the inductive assumption.

As in the proof of Theorem 4.2, we find that $f_{k+1} \in B$. To complete the induction, observe that for $k \geq 2$,

$$\|N_{z^*}^{n_k} f_{k+1} - (z^*)^{n_k}\|_{n_k} \leq \|N_{z^*}^{n_k} f_k - (z^*)^{n_k}\|_{n_k} + \|N_{z^*}^{n_k} h_k\|_{n_k}$$

$$\leq \|N_{z^*}^{n_{k-1}} f_k - (z^*)^{n_{k-1}}\|_{n_{k-1}} + \delta_{n_k} \leq \sum_{j=1}^{k} 2^{-j}.$$

The construction of the sequences $\{f_k\}$, $\{n_k\}$, and z^* is completed.

As in the proof of Theorem 4.2 one can show that there exists $f^* \in B$ such that $\lim_{k\to+\infty} f_k = f^*$, and that

$$\|Sf^* - Sf_k\| \geq \tfrac{1}{2} \cdot \|Sf_{k+1} - Sf_k\|,$$

for $k \geq 1$.

We now prove that $z^* \in Z(f^*)$. For $m > k$, by using (AS.3) we get

$$\|N_{z^*}^{n_k} f_m - (z^*)^{n_k}\|_{n_k} \leq \|N_{z^*}^{n_k} f_{k+1} - (z^*)^{n_k}\|_{n_k} + \sum_{i=k+1}^{m-1} \|N_{z^*}^{n_k}(f_{i+1} - f_i)\|_{n_k}$$

$$\leq \sum_{j=1}^{k} 2^{-j} + \sum_{i=k+1}^{m-1} \|N_{z^*}^{n_i}(f_{i+1} - f_i)\|_{n_i} \leq \sum_{i=1}^{m-1} 2^{-i} \leq 1.$$

Letting $m \to +\infty$ we find that

$$\|N_{z^*}^{n_k} f^* - (z^*)^{n_k}\|_{n_k} \leq 1$$

for $k \geq 1$. This and (AS.3) yield

$$\|N_{z^*}^{n} f^* - (z^*)^{n}\|_{n} \leq 1,$$

$n = 1, 2, \ldots$, as claimed.

Finally, we get

$$\|Sf^* - \Phi^{n_k}((z^*)^{n_k})\| \geq \|Sf^* - Sf_k\| - \|Sf_k - \Phi^{n_k}((z^*)^{n_k})\|$$

$$\geq \tfrac{1}{2} \cdot \|Sf_{k+1} - Sf_k\| - \tfrac{1}{10} \cdot \delta_{n_k} d_{n_k}(N_{z^*}) \geq \tfrac{1}{40} \cdot \delta_{n_k} d_{n_k}(N_{z^*}),$$

$k = 1, 2, \ldots$, which contradicts the fact that $f^* \in B$. This proves that the set $\tilde{A}_4$ and consequently also A_4 are boundary. ∎

Thus, even though a more restrictive concept of perturbed information is considered, we have shown a lower bound corresponding to that from Section 6.

The results of Sections 5 and 6 concerning optimal nonadaptive information and the problem of adaption vs nonadaption can be carried over to the present model (after obvious modifications related to a different concept of the error).

8. Concluding Remarks

In this paper we have considered how noise in information influences the speed of convergence of algorithms for solving linear problems. We assumed that the bounds on the noise are given in advance. This assumption is reasonable as long as we analyze the minimal *error* of algorithms. If we, however, want to consider the minimal *cost* of solving the problem, then it is sometimes appropriate to assume that precision of computing information functionals is subject to change. For instance, suppose that we want to find an ε-approximation ($\varepsilon > 0$) to a function f, based on the values $f(x_i)$, $i = 1, 2, \ldots$. We can compute m_i bits of $f(x_i)$, i.e., a number z_i such that $|z_i - f(x_i)| \leq 2^{-m_i}$, where $0 \leq m_i \leq +\infty$. The cost of computing z_i grows with m_i.

We wish to calculate an ε-approximation with minimal total cost (measured by the time or memory needed). Then it is not reasonable to compute $f(x_i)$ with excessive precision, since the total cost grows with m_i. On the other hand, all z_i must be sufficiently close to $f(x_i)$, since otherwise we may loose the speed of convergence (as shown by the results of this paper), which once more increases the total cost. Thus the problem of finding the best numbers m_i emerges, i.e., we must find the most appropriate noise levels with which information is to be computed. The precise formulation of these issues and the results in this direction will be reported in the future.

References

Chaloner, K. (1984), Optimal Bayesian experimental designs for linear models, *Ann. Statist.* **12,** 283–300.

Kacewicz, B. Z. (1987a), Asymptotic error of algorithms for solving nonlinear problems, *J. Complexity* **3,** 41–56.

Kacewicz, B. Z. (1987b), Optimal solution of ordinary differential equations, *J. Complexity* **3,** 451–465.

Kacewicz, B. Z., Milanese, M., Tempo, R., and Vicino, A. (1986), Optimality of central and projection algorithms for bounded uncertainty, *Systems Control Lett.* **8,** 161–171.

Kadane, J. B., Wasilkowski, G. W., and Woźniakowski, H. (1988), On adaption with noisy information, *J. Complexity* **4,** 257–276.

Kimeldorf, G. S., and Wahba, G. (1970), A correspondence between Bayesian estimation of stochastic processes and smoothing by splines, *Ann. Math. Statist.* **41,** 495–502.

LEE, D., PAVLIDIS, T., AND WASILKOWSKI, G. W. (1987), A note on the trade-off between sampling and quantization in signal processing, *J. Complexity* **3,** 359–371.

MARCHUK, A. G., AND OSIPENKO, K. Y. (1975), Best approximation of functions specified with an error at a finite number of points, *Math. Notes* **17,** 207–212.

MELKMAN, A. A., AND MICCHELLI, C. A. (1979), Optimal estimation of linear operators in Hilbert spaces from inaccurate data, *SIAM J. Numer. Anal.* **16,** 87–105.

MICCHELLI, C. A., AND RIVLIN, T. J. (1977), A survey of optimal recovery, *in* "Estimation in Approximation Theory," pp. 1–54, Plenum, New York.

PLASKOTA, L. (1989), Asymptotic setting for the global maximum of functions in s-dimensions, *J. Complexity* **5,** 369–378.

PLASKOTA, L. (1990), On average case complexity of linear problems with noisy information, *J. Complexity,* **6,** 199–230.

SIKORSKI, K., AND TROJAN, G. M. (1987), Asymptotic optimality of the bisection method, *Numer. Math.,* to appear.

TRAUB, J. F., WASILKOWSKI, G. W., AND WOŹNIAKOWSKI, H. (1988), "Information-Based Complexity," Academic Press, New York.

TROJAN, G. M. (1983), Asymptotic setting for linear problems (see Traub, J. F., Wasilkowski, G. W., and Woźniakowski, H. (1988), "Information-Based Complexity," Academic Press, New York, pp. 383–295).

TROJAN, G. M. (1985), Asymptotic setting for linear problems with nonexact information, unpublished manuscript.

WAHBA, G. (1975), Smoothing noisy data by spline functions, *Numer. Math.* **24,** 383–393.

WAHBA, G. (1984), Cross-validated spline methods for the estimation of multivariate functions from data on functionals, *in* "Statistics: An Appraisal," Proc. 50th Anniversary Conference, Iowa State Stat. Laboratory, pp. 205–235, The Iowa State Univ. Press, Ames, IA.

WASILKOWSKI, G. W., AND WOŹNIAKOWSKI, H. (1987), On optimal algorithms in an asymptotic model with Gaussian measure, *SIAM J. Math. Anal.* **3,** 632–647.